Applied Microeconometrics Problem Set 2

Tanya Rajan

December 12, 2020

Question 1. Suppose that we observe a repeated cross-section of individuals i over two time periods. Let $T_i \in \{1, 2\}$ denote the time period in which we observe individual i. We are interested in the effect of a binary treatment $D_i \in \{0, 1\}$ on an outcome Y_i. Let $Y_i(0)$ and $Y_i(1)$ denote potential outcomes, and assume that there are constant treatment effects, so that $Y_i(1) - Y_i(0) = \alpha$ is a non-stochastic scalar. Suppose that each individual i is a member of a binary group $G_i \in \{0, 1\}$. Maintain the common trends assumption that

$$\mathbb{E}\left[Y_i(0) \mid T_i = 2, G_i = g\right] - \mathbb{E}\left[Y_i(0) \mid T_i = 1, G_i = g\right]$$

does not depend on $g = 1, 2$. In contrast to the usual difference-in-differences design, suppose that some individuals in both groups are treated in both periods, but that the increase between time periods 1 and 2 of the proportion treated in group 1 is larger than the same increase in group 0 . Is α point-identified? If so, propose an estimator that could be used to estimate α. If not, provide a counter-example.

First define $p(g, t) = \mathbb{P}(D_i = 1|G_i = g, T_i = t)$. We are given that $p(1, 2) - p(1, 1) \geqslant p(0, 2) - p(0, 1)$. We observe the following quantities for each combination of group and time:

$$\begin{aligned}
\mathbb{E}[y \mid G = g, T = t] &= p(g, t)\mathbb{E}[Y(1) \mid G = g, T = t] + (1 - p(g, t))\mathbb{E}[Y(0) \mid G = g, T = t] \\
&= \mathbb{E}[Y(0) \mid G = g, T = t] + \alpha p(g, t)
\end{aligned}$$

Then we can subtract off time trends within each group.

$$\begin{aligned}
\Delta_g \equiv\ & \mathbb{E}[Y \mid G = g, T = 2] - \mathbb{E}[Y \mid G = g, T = 1] \\
& = \mathbb{E}[Y(0) \mid G = g, T = 2] - \mathbb{E}[Y(0) \mid G = g, T = 1] + \alpha[p(g, 2) - p(g, 1)]
\end{aligned}$$

Note that by the common trends assumption we know that $\mathbb{E}[Y(0) \mid G = g, T = 2] - \mathbb{E}[Y(0) \mid G = g, T = 1]$ is the same regardless of group g. So subtracting the time-differenced quantities of each group, these terms cancel and we are left with:

$$\Delta_1 - \Delta_0 = \alpha[p(1, 2) - p(1, 1) - p(0, 1) + p(0, 0)]$$

We know that the term inside of brackets is nonzero since we are given that $p(1, 2) - p(1, 1) \geqslant p(0, 2) - p(0, 1)$. So an estimator that point-identifies α (assuming we observe G, T, D) is:

$$\frac{\Delta_1 - \Delta_0}{p(1, 2) - p(1, 1) - p(0, 1) + p(0, 0)}$$

Question 2. Consider the following data generating process for panel data with times $t = 1, \ldots, 5$ and individuals $i = 1, \ldots, n$:

$$\begin{aligned} E_i &\sim \text{Unif}\{2, \ldots, 5\} \\ Y_{it}(0) &= -.2 + .5E_i + U_{it} \\ Y_{it}(1) &= -.2 + .5E_i + \sin(t - \theta E_i) + U_{it} + V_{it} \end{aligned}$$

where V_{it} is serially independent standard normal, independent of E_i, and U_{it} follows the autoregressive process

$$U_{it} = \rho U_{i(t-1)} + \epsilon_{it} \text{ for } t = 2, \ldots, 5 \text{ with } U_{i1} = \epsilon_{i1}$$

where $\epsilon_{it} \sim N(0, 1)$ for each $t = 1, \ldots, 5$, independently of both E_i and V_{it} Note: Angles are measured in radians for the sine function.

(a) Does common trends hold in this data generating process?

The common trends assumption for this event study design requires that we can proxy the unobserved $\mathbb{E}[Y_i(0)|E_i = e]$ as follows:

$$\mathbb{E}[Y_i(0) \mid E_i = e] = \mathbb{E}[Y_{i1}(0) \mid E_i = e] + E\left[Y_{it}(0) - Y_{i1}(0) \mid E_i = e'\right], \text{ for } e' > t, t \geq e$$

This is equivalent to saying that we assume:

$$\mathbb{E}[Y_{it}(0) - Y_{i_1}(0) \mid E_i = e] = \mathbb{E}\left[Y_{it}(0) - Y_{i1}(0) \mid E_i = e'\right], \text{ for } e' > t, t \geq e$$

Evaluating this expression for our given DGP, we get:

$$\mathbb{E}[Y_{it}(0) - Y_{i},(0) \mid E_i = e] = \mathbb{E}[-0.2 + 0.5e + U_{it} + 0.2 - 0.5e - U_{i1} \mid E_i = e] = \mathbb{E}[U_{it} - U_{i1} \mid E_i = e]$$
$$\mathbb{E}\left[Y_{it}(0) - Y_{i1}(0) \mid E_i = e'\right] = \mathbb{E}\left[-0.2 + 0.5e' + U_{it} + 0.2 - 0.5e' - U_{i1} \mid E_i = e\right] = \mathbb{E}[U_{it} - U_{i1} \mid E_i = e]$$

Evaluating U_{it}, we can write that:

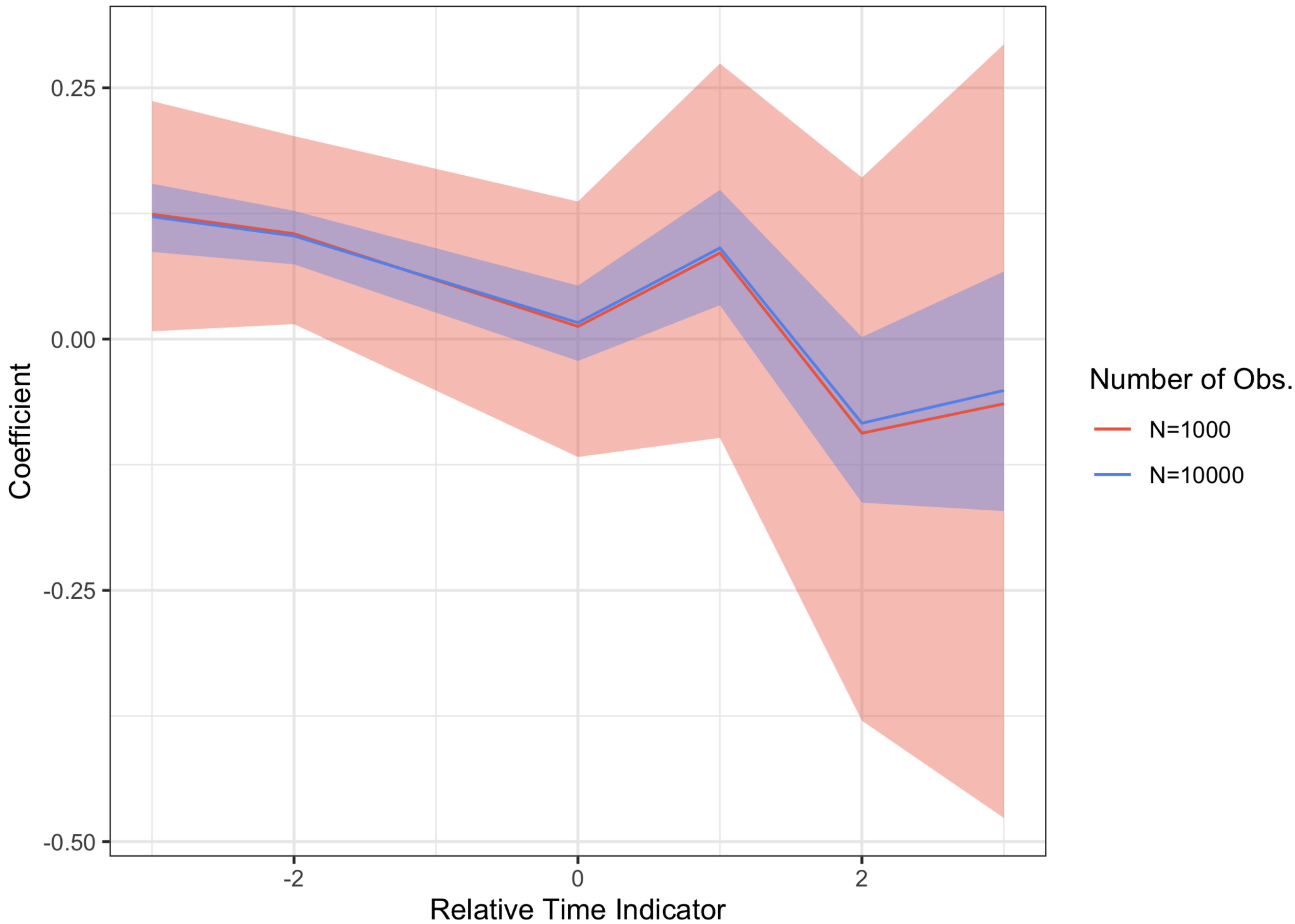
$$U_{it} = \sum_{j=0}^{t-1} \rho^j \varepsilon_{i(t-j)}$$

Since ε_{it} is independent of E_i, we know that $\mathbb{E}[U_{it} - U_{i1}| \mid E_i = e] = \mathbb{E}[U_{it} - U_{i1}]$. So the common trends assumption should hold.

(b) Let $\theta = -2$ and $\rho = .5$ Consider a regression of Y_{it} onto a full set of cohort fixed effects, a full set of time fixed effects, and relative time dummies $D_{it}^r \equiv 1[t - E_i = r]$ for all $r \in \{-4, \ldots, 3\}$ except $r = -1$ and $r = -4$. Run a Monte Carlo simulation with $n = 1000$ and $n = 10000$ to evaluate the finite sample distribution for the coefficients on the relative time dummies. Report the mean together with the 2.5% and 97.5% quantiles using a single figure that combines both sample sizes and has relative time on the horizontal axis.

The figure below shows the coefficients on the relative time indicators (omitting $r = -1$, $r = -4$) for different sample sizes. I run a Monte Carlo simulation with $M = 500$ draws to compute the figures below.

Figure 1: Coefficients on Relative Time Indicators, $\theta = -2$



(c) Repeat the previous part with $\theta = 0$ and $\theta = 1$. Compare your findings across the three different values of θ. Explain any differences, and discuss implications for empirical practice.

The figures below plot the same graph as above for different values of θ. The parameter θ determines the shift in the sin function for each cohort. For $\theta = 0$, cohort does not matter at all—in fact, all cohorts experience the same relative time effects. Eyeballing these graphs, we might be tempted to say that common trends does not hold when $\theta = -2$, even though we know that it does given the DGP we simulated. Similarly, both of the specifications below do pass the eyeball test but give very different estimates of the coefficients on relative time indicators.

Figure 2: Coefficients on Relative Time Indicators, $\theta = 0$

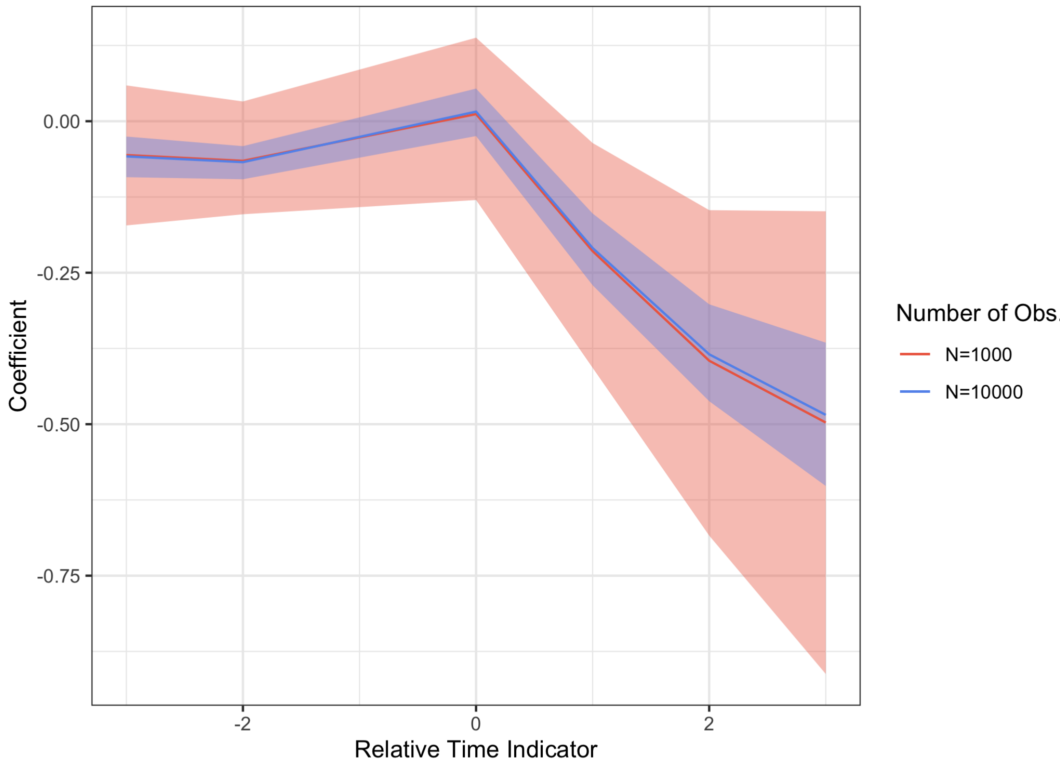
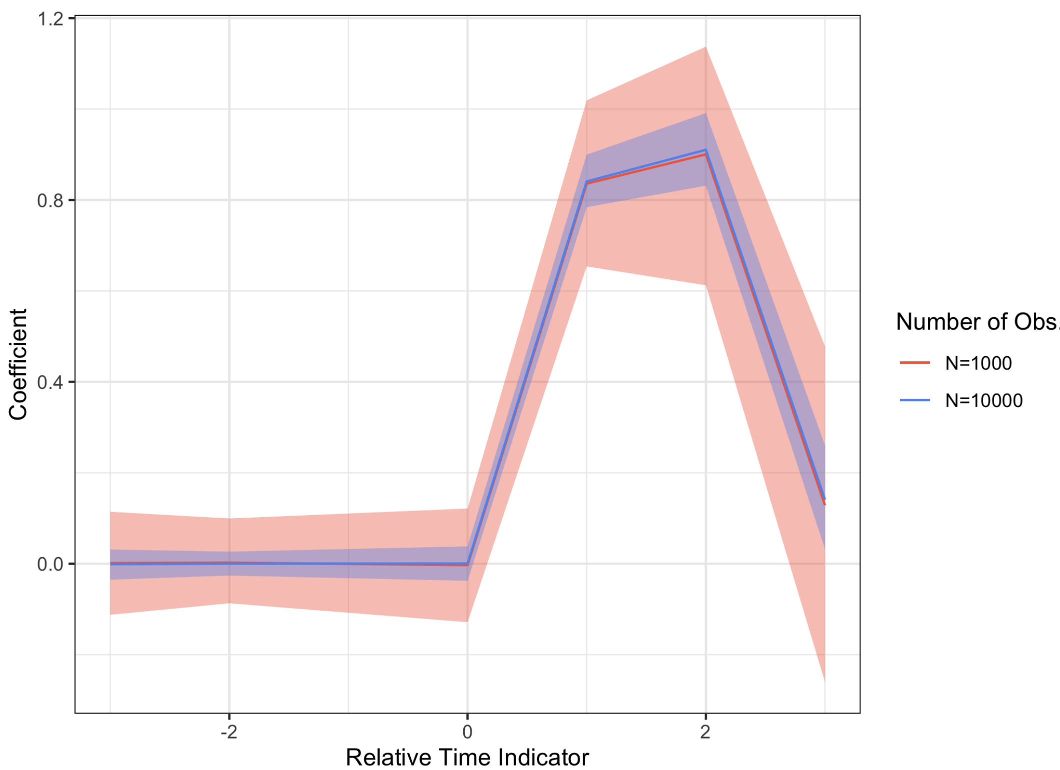


Figure 3: Coefficients on Relative Time Indicators, $\theta = 1$



(d) Devise a consistent estimator of

$$\text{ATE}_3(2) \equiv \mathbf{E}\left[Y_{i3}(1) - Y_{i3}(0) \mid E_i = 2\right]$$

Verify that your estimator works using a Monte Carlo with $n = 10000$ for every value $\theta \in \{-2, 0, 1\}$ and $\rho = .5$

This is simply an extension of what I laid out in the common trends assumption of part (a). The ATE can be written as:

$$\text{ATE}_3(2) = \underbrace{\mathbb{E}\left[Y_{i3}(1) \mid E_i = 2\right]}_{\text{observed}} - \underbrace{\mathbb{E}\left[Y_{i3}(0) \mid E_i = 2\right]}_{\text{unobserved}}$$

We can identify the observed term from the data. Then what we want to identify is the remaining unobserrved term. As stated in part (a), we can rewrite this (using the common trends assumption) as:

$$\mathbb{E}\left[Y_{i3}(0) \mid E_i = 2\right] = \mathbb{E}\left[Y_{i1}(0) \mid E_i = 2\right] + E\left[Y_{i3}(0) - Y_{i1}(0) \mid E_i = 4\right]$$

Everything on the RHS of this equation is observed. Note that I chose cohort 4 to identify this, but any cohort such that $e' > 3$. Using the sample analogues to each of the expectations found above, we have a consistent estimator of $\widehat{ATE}_3(2)$. Below I present ATEs estimated using the proposed estimator and true ATEs calculated using the DGP and unobserved otucomes.

Table 1: Estimated vs. Actual ATEs for Various Values of θ

	Actual	Estimated
$\theta = -2$	0.6606	0.6584
$\theta = 0$	0.1412	0.1398
$\theta = 1$	0.8416	0.8401

(e) For this part let $\theta = 1$ Consider the same regression as in part b). Run a Monte Carlo for each combination of $n = 10, 20, 50, 200$ and $\rho = 0, .5, 1$ that evaluates the empirical size (rejection probability when the null hypothesis is true) of a level $.05t$ -test for the coefficient on D_{it}^1 implemented using each of the following approaches:

- The classical asymptotic variance estimator under homoskedasticity.
- An Eicker-Huber-White heteroskedasticity-robust asymptotic variance estimator. (Use the HC(1) version.)
- The cluster-robust asymptotic variance estimator, clustering over individuals i. Use the finite-sample correction discussed in the supplemental notes.
- The clustered Wild bootstrap, clustering over individuals i.

Report your results using a single, well-designed table and discuss the relative performance of the different approaches. Note: You can keep E_i as deterministic across simulations, with an equal number of observations for each value of E_i. If you don't do that, then you might have simulation draws in which the regression coefficients do not exist due to perfect collinearity.

I first derive the true value of β_{ev}^1 under the null hypothesis. From the supplemental notes, we know that when common trends holds and when we assume that the ATE is the same across cohorts (which is true when $\theta = 1$ as discussed above), then we can write,

$$[\beta_{\text{ev}}]_j = \sum_{r \in \mathcal{R}_j} \text{ATE}_r - \sum_{r \in \mathcal{R}_0} \text{ATE}_r$$

So we have to subtract the ATE when relative time is equal to 0 from the ATE when relative time is equal to 1. Considering our DGP equations, this gives us: $\sin(1) - \sin(0) = \sin(1)$ as the true value. Note that this interpretation works because we only omitted relative time indicators *pre-treatment*. Otherwise we would have to interpret β^1_{ev} as the treatment effect relative to the treatment effect in the omitted post-treatment relative time bin.

The table below presents the results from testing this null hypothesis using various methods. I run a Monte Carlo simulation with $M = 500$ draws. I was having issues inverting matrices with sample size $N = 20$, so I had to raise the smallest sample size to $N = 40$. In general, homoskedastic standard errors seem to return empirical rejection rates lower than expected since they do not account for autocorrelation in error terms. Robust standard errors seem to perform better, but also are overconfident when ρ becomes high. The clustered robust errors (clustered at the individual level) perform very well for all values of ρ in large samples since they allow for dependencies in error terms within individuals but across time. I realized rather late that I clustered the bootstrap by time instead of individual. I did not have time to rerun the estimation on the whole sample. For this particular test, I ran a simulation with $M = 100$ draws and $B = 50$ bootstraps. The results show that the wild bootstrap has the advantages of the Clustered Robust SEs since it accounts for serial correlation and it performs well in small sample.

Table 2: Empirical Rejection Rates Using Various Methods to Compute SEs

	Sample Sizes		
	N=40	N=50	N=200
Homoskedastic SEs			
$\rho = 0$	0.030	0.042	0.032
$\rho = 0.5$	0.030	0.030	0.024
$\rho = 1$	0.012	0.008	0.010
Robust SEs			
$\rho = 0$	0.046	0.058	0.054
$\rho = 0.5$	0.040	0.046	0.046
$\rho = 1$	0.030	0.024	0.018
Clustered Robust SEs			
$\rho = 0$	0.044	0.072	0.050
$\rho = 0.5$	0.056	0.060	0.050
$\rho = 1$	0.070	0.062	0.054
Clustered Wild Bootstrap			
$\rho = 0$	0.050	0.040	0.050
$\rho = 0.5$	0.050	0.050	0.040
$\rho = 1$	0.060	0.040	0.050

This table reports empirical rejection rates in a Monte Carlo simulation of $M = 500$ draws. Errors are clustered at the individual level. Bootstraps were run on 50 replications.

Question 3. Replicate Figures 10 and 11 in "Combining Matching and Synthetic Control to Trade off Biases from Extrapolation and Interpolation" by Kellogg, Mogstad, Pouliot, and Torgovitsky (September, 2020). The data is originally from "The Economic costs of Conflict: A Case Study of the Basque Country" by Abadie and Gardeazabal (2003), and is available on Canvas.

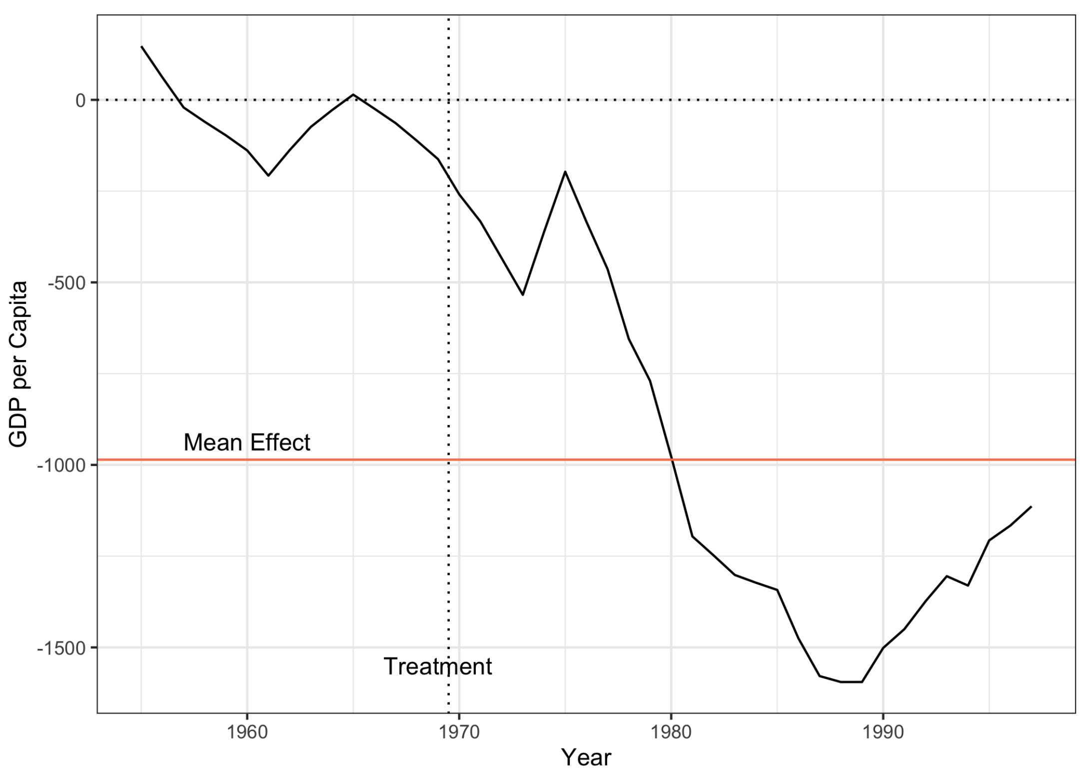
The figures below are the replication. I code up synthetic control and matching estimators and then combine them into the MASC estimator:

$$\gamma^{MASC} = \phi\gamma_k^{MA} + (1-\phi)\gamma^{SC}$$

where k represents the number of nearest neighbors to match with. In order to find the appropriate value for ϕ and for k, I create a grid of values for k between 1 and 10. For fixed k, I then crossvalidate the MASC estimator to find the optimal value for ϕ. Using 7 folds (years 1960 - 1967, I train the MASC estimator on years $t \in [1955, \bar{t}_F]$ and then run the validation on outcomes in year $\bar{t}_f + 1$. I calculate the average error over all folds, as well as the optimal value of ϕ from the closed form expression provided in the paper:

$$\hat{\phi}(m) \equiv \frac{\sum_{f=1}^{F}\left(\hat{\gamma}_f^{\text{ma}}(m) - \hat{\gamma}_f^{\text{sc}}\right)\left(y_{1,\bar{t}_f+1} - \hat{\gamma}_f^{\text{sc}}\right)}{\sum_{f=1}^{F}\left(\hat{\gamma}_f^{\text{ma}}(m) - \hat{\gamma}_f^{\text{sc}}\right)^2}$$

Figure 4: Pre-terrorism Fit and Post-Period Estimated Costs of Terorrism using MASC



I present my Figure 11 replication below. It is close but not exactly the same as the figure presented in the paper. I think one issue may be that I used an optimizer where equality constraints are not easy to include—so I had to give it some tolerance level for the constraint (e.g. weights may sum up to 1.001). It seems that there is something off about the magnitudes that my estimators estimate as well. However, given that Figure 10 matches the paper exactly, I believe the issue may be with the penalized SC estimator.

Figure 5: Difference from MASC in Cost of Terrorism